

## Den Danske Ordbog som begrebsordbog

This article describes how wordnet information can be used to enrich semantic information in an electronic dictionary. The semantic division and definitions from The Danish Dictionary were used to build the Danish wordnet, *DanNet*, so there is a close connection between the two resources already. Especially for language production purposes and in second language learning, human users can benefit from the onomasiological organization principle of a wordnet. In the online version of The Danish Dictionary the wordnet information is presented in a new element, called related words, which puts together not only synonyms but a wider range of paradigmatic alternatives, including hyponyms and hyperonyms, that are relevant for someone who is looking for “the right word” when writing a text, or for someone trying to learn the vocabulary of a particular topic. The principles of selection and presentation are described, with particular attention devoted to the problems and the challenges that remain for future development.

Nøgleord: wordnet, onomasiologisk søgning, hypo-/ hyperonym, begrebsordbog, synonymordbog, taksonomi, ontologi

En begrebsordbog er nyttig for den der ønsker at udvikle sine kommunikative færdigheder til at udtrykke sig kreativt, præcist og nuanceret, på skrift og i tale. Den minder om synonymordbogen, men dækker ofte bredere ved at inddrage flere betydningsrelationer end synonymi, navnlig hyperonymer (ord der udtrykker overbegreber), hyponymer (ord der udtrykker underbegreber) og undertiden antonymer. Og til forskel fra betydningsordbogens semasiologiske ordningsprincip er tilgangen til lemmaerne i begrebsordbogen onomasiologisk, dvs. ordnet efter deres betydning. Den emnemæssige opdeling gør også begrebsordbogen nyttig i indlæringsmæssig sammenhæng, både for L2-studerende og for den der ønsker at sætte sig systematisk ind i sprogets byggesten, ordforrådet.

En betydningsordbog kan altså godt indeholde præcis de samme lemmaer som en begrebsordbog. Men den trykte ordbog kan ikke rumme begge typer tilgang samtidig. Det skyldes at bogens sider kun har to dimensioner, og i en trykt ordbog må man derfor vælge hvilket princip der følges. E-ordbogen er ikke underlagt den samme begrænsning. Muligheden for ved hjælp af hyperlink at klikke sig fra én ordbogsartikel til en anden åbner for en anden tilgang til ordbogens indhold, fx til semantisk beslægtede ord og betydninger. Lige fra de første planer for en netudgave af Den Danske Ordbog (herefter DDO) blev udtænkt, har denne mulighed været højt prioriteret af redaktionen. Ønsket har været at etablere klikbare sammenhænge mellem ord og betydninger som i forvejen er indeholdt i ordbogen, men som på grund af den alfabetiske tilgangsstruktur ikke umiddelbart er forbundet

med hinanden i den trykte ordbog. Nogle af disse tanker har også været publiceret tidligere (fx Trap-Jensen 2005a), men først nu hvor ordbogen er realiseret i en netversion, kan der gøres status: Hvilke tanker ligger bag, hvordan er ønskerne blevet løst, og hvilke overvejelser og valg har skullet træffes undervejs.

## 1. Udgangspunkt og grundlæggende antagelser

Allerede under tilblivelsen af den trykte ordbog blev det onomasiologiske perspektiv indtænkt på forskellig vis i det redaktionelle arbejde. Muligheden for at udtrække et hyponymihierarki blev lettet af at der til alle definitioner blev knyttet et særligt indholdselement hvori redaktørerne anførte *genus proximum* for den pågældende betydning. På den måde blev det nærmeste overbegreb indholdsopmærket til senere brug. Noget tilsvarende skete for angivelse af fagområde, idet der blev indført et element til angivelse af en betydnings faglige placering hvis en sådan på nogen måde kunne specificeres. Ord som *skole*, *kridt* og *tavle* er en del af det almene ordforråd og er derfor naturligvis ikke blevet markeret med en fagbetegnelse i ordbogen, men alligevel er det registreret at (en bestemt betydning af) ordene hører til det faglige område ‘undervisning’. Også denne oplysning kan udnyttes i forbindelse med en systematisk betydningsklassifikation.

Indførelsen af disse elementer i artikelstrukturen var ganske fremsynet, men heller ikke uden problemer. Netop fordi den fremtidige udnyttelse kun var antydningvis formuleret, var de redaktionelle principper for deres udfyldelse, naturligt nok, tilsvarende ufuldkomne. En tidlig analyse af *genus proximum*-feltet bragte også visse inkonsekvenser for dagen: Redaktørerne har ikke været enige om valg af overbegreb for den ikke-taksonomiske del af ordforrådet: Hvad er overbegrebet for abstrakter, fx verbalsubstantiver? Mange verbalsubstantiver følger skabelonen: ‘det at gøre X’, og her har redaktørerne undertiden valgt det indholdstomme *det*, (*afrusning*: ‘det at gøre en beruset person ædru’), andre gange den verbale handling (*effektivisering*: ‘det at gøre noget mere effektivt’) og andre gange igen det mere indholdstunge objekt for verbalhandlingen (*rejsning*: ‘det at gøre oprør mod noget’). En tilsvarende usikkerhed har gjort sig gældende for mange adjektiver og adverbier samt for de lukkede ordklasser. Det er naturligvis ikke så underligt eftersom de nævnte typer ikke forklares intensionelt efter den aristoteliske model. Denne finder især anvendelse på betydninger inden for et taksonomisk hierarki.

For den taksonomiske del af ordforrådet var der det problem at ordbogens definitioner primært blev udformet så de fungerede på stedet for den menneskelige bruger. Der blev ikke

foretaget systematiske tjek på tværs af artiklerne for at sikre at den systematiske klassifikation blev overholdt. Når man bagefter foretager et sådant tjek, viser der sig derfor også tilfældige uoverensstemmelser i systemet. En *kanin* er fx defineret som en gnaver, en *hare* derimod som et pattedyr. Og en *zebra* er defineret som et hovdyr, et *muldyr* som et dyr. Den manglende konsistens er endnu mere udtalt for laverestående organismer som skive-, sækspore- og stilksporesvampe eller brun-, grøn- og rødalger. De uoverensstemmelser der kan påvises, gør ikke den enkelte definition forkert, og et krav om at definitionerne skal følge den faglige klassifikation og indeholde nærmeste taksonomiske overbegreb, ville føre til alt for komplicerede faglige forklaringer, ikke mindst for ord af den sidstnævnte type. Men det betyder at et automatisk genereret hierarki udelukkende baseret på udtræk fra ordbogens genus proximum-elementer ikke er stringent og systematisk. I et automatisk genereret hierarki vil kategorien *gnavere* omfatte *kanin*, men ikke *hare*, og kategorien *hovdyr* vil omfatte *zebra*, men ikke *muldyr*.

Problemer af denne karakter førte til at redaktionen besluttede sig for en mere systematisk, tværgående opmærkning af ordbogens betydninger. Resultatet af overvejelserne førte til at Det Danske Sprog- og Litteraturselskab engagerede sig i projektet DanNet i et samarbejde med Center for Sprogteknologi ved Københavns Universitet. Heri er ikke blot over-/underbegrebshierarkiet opmærket, men tillige en række andre semantiske relationer mellem ordbogens ord og betydninger. DanNet følger modellen kendt fra Princeton WordNet og EuroWordNet og har en række selvstændige sprogteknologiske anvendelsesmuligheder. For redaktionen af DDO har et vigtigt perspektiv været at bruge DanNet til en forbedret onomasiologisk datatilgang i netordbogen.

## 2. Beslægtede ord i Den Danske Ordbog

I den første netversion af DDO bruges oplysninger fra DanNet til en ny funktion: "Beslægtede ord". Under denne overskrift kan brugeren finde ord der kan stå på samme paradigmatiske plads som opslagsordet i den pågældende betydning, dvs. ord der især er nyttige ved tekstproduktion til "at finde det rette ord", og som desuden kan være oplysende for ikke-modersmålstalende og andre der er interesseret i at lære de ord der hører til et bestemt betydningsområde.

Brugeren får en række forslag til alternative ord der kan bruges ved skrivning af tekster. Ved tekstproduktion har man ofte brug for at variere sit ordvalg, men de alternative muligheder begrænser sig ikke til synonymer. I nogle sammenhænge kan et mere generelt ord

være på sin plads, i andre tilfælde er det nærliggende at bruge et mere specifikt. I konsekvens heraf omfatter oplysningstypen “Beslægtede ord” både hyperonymer, hyponymer og ko-hyponymer. Et eksempel er vist i figur 1.

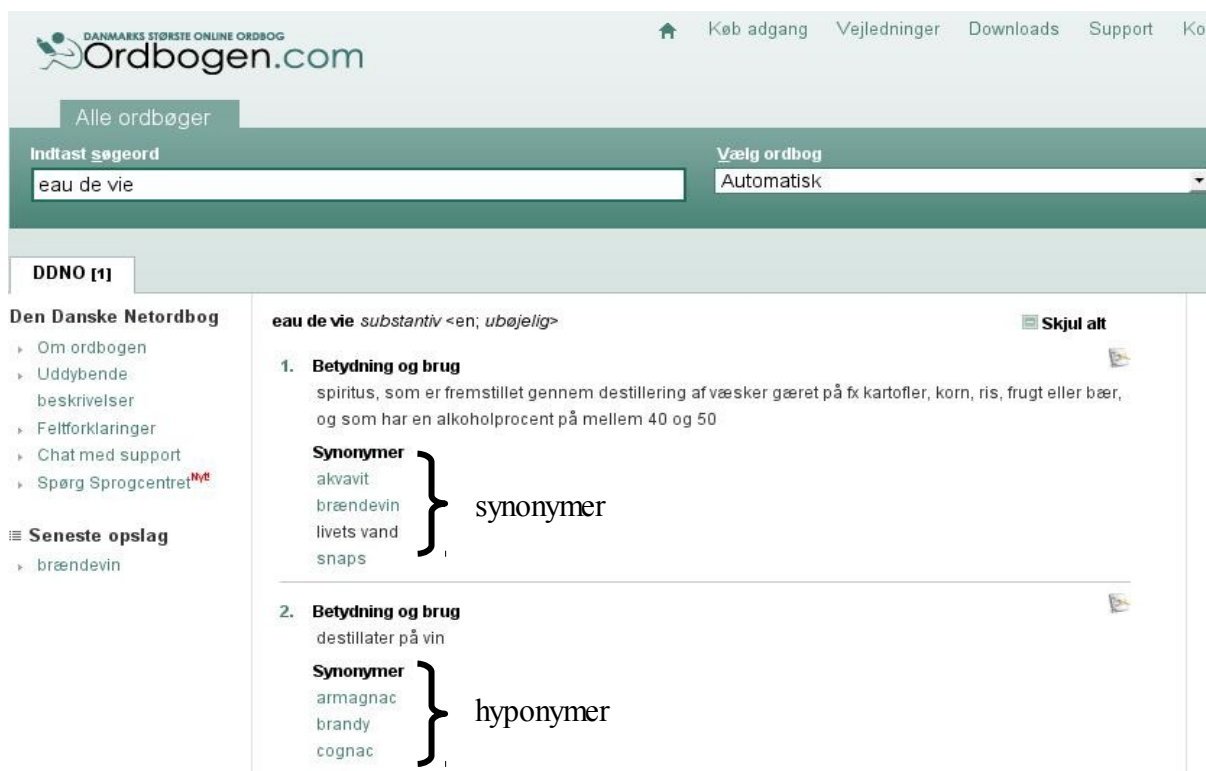
Denne datatype kan ses som en slags udløber af den diskussion der fandt sted i forbindelse med anmeldelserne af DDO i LexicoNordica (Bergenholtz og Vrang 2005: 180f., Trap-Jensen 2005b: 189ff., Bergenholtz og Vrang 2006: 191 ff.). Anmelderne fandt heri at DDO’s synonymudvalg var for snævert til brug for tekstproduktion. Den opfattelse er til dels blevet imødekommet med funktionen “Beslægtede ord”.

The screenshot shows the DDO entry for 'spise'. The main definition is: '1. tilføre kroppen (nødvendig) næring gennem mund og hals; indtage mad'. Below this, there is a section for 'BESLÆGTEDE ORD' which lists numerous related terms such as 'indtage', 'trøstespise', 'delikaterere sig', 'dinere', 'forpise sig', 'foræde sig', 'forsluge sig', 'fråse', 'fylde', 'få 'ned', 'gaffe i sig', 'gnaske', 'gnave', 'guffe', 'gumle', 'have', 'hugge 'i sig', 'klemme 'ned', 'kværne', 'mække sig', 'nyde', 'omsætte', 'proppe', 'putte', 'røre', 'skaffe', 'skovle', 'slikke', 'sluge', 'smovse', 'souperer', 'spise', 'spise op', 'spise sig (i)gennem noget', 'stikke', 'tømme', 'æde', 'æde 'op', 'æde sig (i)gennem', 'æde sig noget til', 'fortære', 'nappe', 'spise ude', and 'andre ord med "indtage" som overbegreb: drikke, konsumere, ryge, sluge, snuppe, slubre, kyle, høvle, hælde'. To the right, a search results sidebar shows a list of related words including 'spiritusproblem sb.', 'spiritusprøve<sup>1</sup> sb.', 'spiritusprøve<sup>2</sup> vb.', 'spirituspåvirket adj.', 'spiritusånde sb.', 'spirituøs adj.', 'spirometer sb.', 'spirrevip sb.', 'spiræa sb.', 'spise<sup>1</sup> sb.', and 'spise<sup>2</sup> vb.', with 'spise<sup>2</sup> vb.' being the selected entry.

Figur 1. Beslægtede ord for verbet spise i DDO

Redaktionen har dog ikke ønsket at udvide synonymibegrebet til også at omfatte over-, under- og søsterbegreber sådan som det gøres i fx Den Danske Netordbog. Et enkelt eksempel herfra kan tjene til sammenligning. Den Danske Netordbog anfører to betydninger af lemmaet *eau de vie* som vist i figur 2. Den første har definitionen “spiritus, som er fremstillet gennem

destillering af væsker gæret på fx kartofler, korn, ris, frugt eller bær, og som har en alkoholprocent på mellem 40 og 50”, og hertil gives synonymerne *akvavit*, *brændevin*, *livets vand* og *snaps*. Betydning 2 har definitionen “destillater på vin” og synonymerne *armagnac*, *brandy* og *cognac*. De første fire synonymymer er synonymymer i egentlig forstand, dvs. ordene udpeger samme entitet som opslagsordet<sup>1</sup>. De tre sidste er derimod ikke synonymymer, idet definitionen beskriver en fællesbetegnelse (signaleret ved hjælp af pluralisformen), og *armagnac*, *brandy* og *cognac* skal derfor opfattes som eksempler der tilhører fællesbetegnelsen. De er med andre ord hyponymer, ikke synonymymer.



Figur 2: Den Danske Netordbog

<sup>1</sup> Forudsat naturligvis at definitionen godtages. Efter vores mening dækker definitionen ikke ordets brug i almensproget, men det er en anden sag

Modsat Den Danske Netordbog fastholder redaktionen af DDO sin mere snævre opfattelse af synonymymer, og vi ønsker derfor at bevare den trykte ordbogs synonymymer som en selvstændig oplysningstype. Vi mener det er en styrke at brugeren fortsat kan stole på at et synonym er et ord der bruges med samme betydning som opslagsordet, og at oplysningen kan bruges ved både reception og produktion. Men vi anerkender brugerens behov for et bredere udvalg af beslægtede ord til tekstproduktion; blot foretrækker vi at udskille dem som en selvstændig kategori så de semantiske typer ikke blandes sammen. Derfor bringer DDO synonymymer og beslægtede ord som selvstændige, adskilte oplysningstyper.

### 3. Fra DanNet til “Beslægtede ord” i Den Danske Ordbog

Uden at gå i detaljer med baggrunden og principperne for DanNets opbygning (læs fx nærmere om DanNet i Pedersen m.fl. 2009) skal her blot nævnes et par elementære træk ved et wordnet. Grundenheden er ikke ord, men *begreber*, eller mere præcist: grupper bestående af et eller flere synonymymer der udtrykker det samme begrebsindhold. Det samlede ordforråd opdeles i et antal grove semantiske klasser, ca. 200 i alt, i wordnet-verdenen benævnt ‘ontologiske typer’. Eksempler på ontologiske typer i DanNet er *Natural+Substance* (is, lava, sand), *Plant+Object+Comestible* (avocado, gulerod, tomat), *Human+Object+Occupation* (revisor, sygeplejerske, taxachauffør) og *UnboundedEvent+Agentive+Mental* (reflektere, analysere, tænke), her gengivet med den internationale standardkodnings betegnelser og derfor på engelsk. Under kodningen angives for hvert begreb hvilken ontologisk type det tilhører. Den ontologiske type fungerer som en slags skabelon, idet den ontologiske type bestemmer hvilke semantiske relationer det er relevant at kode for et givet begreb. I alt opererer DanNet med 18 forskellige semantiske relationer der kan specificeres for et begreb. De almindeligste relationer er *hyponymi*, *hyperonymi*, *del-helhed*, *antonymi*, *nær-synonymi*, *formål*, *vedrører* og *involveret aktør*.

Som nævnt indeholder den første netudgave af DDO “Beslægtede ord” inden for tre forskellige semantiske kategorier: hyperonymer, hyponymer og ko-hyponymer. I det følgende beskriver vi kort hvordan de hver især udvælges og vises.

#### 3.1. Hyperonymer

Udvælgelsen af hyperonymer er uproblematisk da oplysningen umiddelbart kan hentes fra DanNet, og der vil aldrig optræde ret mange ord på denne liste. Det er dog værd at bemærke at der godt kan optræde mere end ét ord på listen. Det sker i to tilfælde: For det første hvis der er flere ord knyttet til det samme begreb og DanNet angiver denne begrebsgruppe som overbegreb. Et eksempel er ordet *jeep*, hvor det under “Beslægtede ord” oplyses at mere generelle ord er: *bil*, *dyt*, *automobil*, *slæde*, *vogn*, *karet*. Årsagen er at disse ord alle har betydningen ‘bil’ og dermed er seks synonyme udtryk for det samme begreb ‘bil’.

For det andet hvis et ord har mere end ét overbegreb. Det er tilfældet for et ord som *rulleskøjte*, der i DanNet er registreret som underbegreb til to forskellige begrebssæt, dels til begrebssættet {*skotøj*, *fodbeklædning*, *fodtøj*}, dels til {*sportsrekvisit*}. Alle fire ord er derfor nævnt som mere generelle ord.

### 3.2. Ko-hyponymer

Ved ko-hyponymer kan der i sjældne tilfælde forekomme flere tusind ord, og det giver ikke mening at præsentere brugerne for en liste med så mange ord. I den første version nøjes vi med at skære listen ned til i første omgang 30 ord, men med mulighed for at få vist op til 200 ord. Når der er så mange ord på listen, bliver det vigtigt at kunne sortere listen sådan at det er de mest relevante ord der vises – og vises først. Det gør vi ved sortere listen efter en score for lighed med opslagsordet. Følgende faktorer indgår når scoren skal beregnes:

Først sammenlignes opslagsordet i den pågældende betydning med hvert enkelt ko-hyponym. Det første der ses på, er ontologisk type. Jo større lighed der er mellem opslagsordet og ko-hyponymets ontologiske type, jo højere bliver scoren.

Dernæst sammenlignes de forskellige relationer der beskriver opslagsordet og ko-hyponymet. Det resulterer i en højere score at have mange relationer tilfælles, men der behøver ikke være fuldstændig overensstemmelse mellem relationerne. Hvis opslagsordet fx er *benzinbil* og dette er beskrevet med relationen HAR-SOM-DEL “benzinmotor”, ligner dette ord *dieselbil* (HAR-SOM-DEL “dieselmotor”) mere end det ligner *ladvogn* (HAR-SOM-DEL “lad”) fordi *dieselmotor* og *benzinmotor* tilhører samme ontologiske type. Og formelt er både *dieselbil* og *ladvogn* mere lig *benzinbil* end *flugtbil* fordi det ikke er relevant overhovedet at beskrive sidstnævnte ord med HAR-SOM-DEL-relationen; i stedet anvendes BRUGES-TIL-relationen (BRUGES-TIL “at flygte”).

På grundlag af scoren for lighed kan man anføre de beslægtede ord i rækkefølgen: *dieselbil*, *ladvogn*, *flugtbil*. Hvis man skulle vælge kun at vise kun ét af de tre ord, ville det blive *dieselbil* som det ord med den højeste score. Og det er netop hvad vi gerne ville opnå.

I figur 3 kan man se for ordet *frakke* hvordan scoren fungerer. Listen med ko-hyponymer begynder med *bluse*, og de første syv ord (indtil *sweater*) scorer højest, hvorefter scoren falder for de følgende. Interesserede brugere kan få vist scoren i onlineversionen ved at holde musen hen over ordene.

Ideelt set er en forudsætning for en præcis beregning af scoren i alle tilfælde at samtlige begreber er udtømmende beskrevet. Men en udtømmende formel beskrivelse af alle begreber er en resursekrævende opgave, og DanNet-projektet dækker ikke alle dele af sproget 100 %. Fuldkodningen har især været koncentreret om de konkrete substantiver, dvs. genstande og andre fysiske entiteter.

Bemærk desuden at visse dele af sproget vanskeligt eller slet ikke lader sig beskrive formelt. Hvordan skulle man fx udtrykke forskellen på ord som *tiger*, *løve* og *puma* ved hjælp af en formalisme? At beskrive at *tiger* HAR-SOM-DEL "striber" og *løve* BRUGES-TIL<sup>2</sup> "at brøle", bliver hurtigt vilkårligt, og en stum løve og en farveløs tiger er stadig en løve henholdsvis en tiger, hvorimod en benzinbil uden benzinmotor er en logisk umulighed. At sortere ord i disse dele af sproget automatisk efter relevans, er ikke muligt.

BESLÆGTEDE ORD<sup>BETA</sup> mere generelt: klædningsstykke | beklædningsgenstand || mere specifikt: cottoncoat | overfrakke | domino | regnfrakke | kåbe || andre ord med "klædningsstykke - beklædningsgenstand" som overbegreb: bluse | jakke | vest | slag | kappe | trøje | sweater | korselet | badedragt | badetøj | tøjstykke | muffe | klædning | dragt | forklæde | særk | sparkedragt | pashmina | alba | latex | mini | draperi | halsedise | rullekrave | busseronne | hose | livstykke | kaftan | festdragt | heldragt | muffedise | slør | kreation | gamache | trevl | underbukser | overdel | underdel | undertrøje | gevandt | skjorte | korset | kjortel | vante | luffe | hovedbeklædning | kåbe | natkjole | buks | bukser | kjole | strømpebukser | skotøj | fodbeklædning | foddøj | underkjole | las | klud | pjalt | bodystocking | klassiker | bh | strømpe | tørklæde | handske | sjal | nederdel | krave | slips ...skjul vis som grafik (eksternt link) 

<sup>2</sup> Eller hvilken vilkårlig formel relation man end måtte vælge at bruge.



*Figur 3: Kohyponymer for frakke bringes i rækkefølge efter faldende score*

### 3.3. Hyponymer

Den sidste liste er listen over hyponymer. I lighed med listen over ko-hyponymer kan denne liste blive meget lang, men til forskel fra denne bliver hyponymlisten ikke sorteret. Hvis opslagsordet er *bil*, hvilken slags bil er så mest relevant: en *benzinbil* eller en *ladvogn*? Det spørgsmål kan vi ikke svare på (endnu), og de indtil 200 ord der vises, er derfor udvalgt vilkårligt.

Det er selvfølgelig ikke hensigtsmæssigt. Vi håber i fremtidige versioner at blive i stand til at præsentere oplysningerne bedre, fx ved at gruppere hyponymerne i relevante typer så brugerne ikke belemres unødigt med alt for mange ord ad gangen.

## 4. Problemer og diskussion

Man skal være opmærksom på at de ord der anføres under “Beslægtede ord”, er automatisk genererede oplysninger fra en resurse der i første omgang er beregnet til brug inden for sprogteknologi, og de kan være fejlbehæftede. Af samme grund er feltet indtil videre forsynet med et lille “beta” efter overskriften. Kodningen i DanNet er ikke fejlfri, og de fejl der måtte være i resursen, bliver selvfølgelig overført til DDO. Nogle af de problemer vi selv er opmærksomme på, diskuterer vi i det følgende.

### 4.1. Ingen eller for få beslægtede ord

DanNet dækker ikke alle DDO's betydninger. Det betyder at der er mange ord og betydninger som kunne have beslægtede ord anført, men som ikke har fået det af den enkle grund at de ikke er kommet med i DanNet. I efteråret 2009 indeholdt DanNet ca. 50.000 begreber. Til sammenligning indeholder DDO ca. 95.000 opslagsord med ca. 120.000 forskellige betydninger, hvilket betyder at over halvdelen af DDO's betydninger ikke tilbyder brugeren "Beslægtede ord". Til gengæld er udvælgelsesprincippet i DanNet til dels frekvensbaseret, og målet er at alle hovedbetydninger af hyppige konkrete substantiver er beskrevet i resursen. En del af de manglende betydninger vil efterhånden blive tilføjet idet DanNet planlægges udvidet til mindst 60.000 begrebssæt med udgangen af 2010.

#### 4.2. For mange beslægtede ord, relevante eller ikke relevante

Wordnet-ontologien indeholder færre kategorier end de fleste begrebsordbøger, som tidligere nævnt omkring 200. Til sammenligning er der i Rogets *Thesaurus* . 1000 emneområder, og et antal på 800-1000 er det der bruges i mange andre begrebsordbøger, som jo for de flestes vedkommende også står i gæld til Roget, fx Andersen, Bring, Dornseiff og Rosbach. Det er formentlig en del af forklaringen på at DanNets ca. 200 ontologiske typer giver en for grov opdeling af ordforrådet når det skal præsenteres for menneskelige brugere.

Et centralt kendetegn ved DanNet er at det afspejler sprogets egne kategoriseringer af verden. Det er først og fremmest det der adskiller et wordnet fra terminologiens ontologier. Vi har derfor ikke villet indføre begrebskategorier i DanNet som ikke har modsvarende sproglige udtryksenheder, dvs. er ord i sproget. Det anser vi for en styrke ved DanNet med tanke på dets anvendelighed i sprogteknologiske applikationer, som jo er det primære anvendelsesområde for resursen. Men det indebærer – med tanke på den menneskelige bruger – samtidig en udfordring at præsentere data i en overskuelig form.

Problemet er stigende, jo højere op i begrebshierarkiet man kommer. Et ord som *person* har således 3.569 hyponymer i DanNet, og dette tal dækker over en højst uensartet gruppe ord fordi sproget ikke er opbygget på samme stringente måde som ontologier eller faglige taksonomier. Det betyder at der optræder semantisk mere generelle ord (fx *mand*, *kvinde*, *barn*) side om side med meget specifikke ord anbragt direkte under overbegrebet *person* (fx *oldtidslevning*, *asylansøger*, *yuppie drabsmand*). Resultatet er en ret uoverskuelig opremsning af ord som ikke har andet tilfælles end at de alle betegner personer.

I traditionelle begrebsordbøger skelnes der for det første ikke mellem hyperonymer, hyponymer og ko-hyponymer; et emneområde kan godt indeholde ord fra samtlige typer. For det andet er det ikke nødvendigvis et krav at en begrebsklasse skal kunne betegnes med en sproglig udtryksform. Deraf følger at man kan operere med mere overskuelige emneområder, og det er en klar fordel, navnlig i den mere abstrakte ende af begrebshierarkiet. Figur 4 viser et eksempel fra Macmillan, som indeholder en manuelt redigeret tesaurusfunktion.

Det er en fordel at brugeren her kun bliver præsenteret for ord som befinder sig på nogenlunde samme abstraktionsniveau. Det er jo typisk dem der er relevante alternativer for den der leder efter et andet ord i forbindelse med tekstproduktion, hvorimod de ovenfor nævnte eksempelord fra DDO *oldtidslevning*, *asylansøger*, *yuppie drabsmand* aldrig vil være reelle paradigmatiske alternativer.

Omvendt skal man være opmærksom på at den i sig selv prisværdige bestræbelse på at operere med antalsmæssigt overskuelige kategorier ikke altid fører til lige meningsfulde emnekategorier. Den der leder efter alternative ord for *chinchillai* Macmillan, bliver således præsenteret for emnet "Pattedyr i Nord-, Mellem- og Sydamerika", med 27 medlemmer (*alpaca*, *armadillo*, *beaver*, *bobcat*, *caribou*, *chinchilla*, *cougar*, *coyote*, *coypu*, *gopher*, *grizzly bear*, *groundhog*, *jackrabbit*, *llama*, *mountain goat*, *mountain lion*, *muskrat*, *pack rat*, *prairie dog*, *puma*, *raccoon/racoon*, *skunk*, *sloth*, *snapping turtle*, *vicuna*, *wapiti*, *woodchuck*), men det er nok sandsynligt at andre gnavere eller andre kæledyr i mange tilfælde vil være bedre paradigmatiske alternativer til *chinchilla* end fx *coyote* (prærieulv), *grizzly* (gråbjørn) eller *caribou* (rensdyr).

Det bliver med sikkerhed en vigtig opgave at udvikle et system der tillader først en hensigtsmæssig gruppering af hyponymer, dernæst en prioritering blandt grupperingerne således at brugerne kan vælge mellem et overskueligt antal relevante kandidater. Det kan næppe lade sig gøre uden en vis mængde manuel redigering, men på den anden side vil vi helst ikke give køb på princippet om kun at operere med sproginterne begrebskategorier. Én metode kunne være at analysere den formelle beskrivelse i DanNet nærmere og gruppere

thesaurus entry for **person**[back to definition of person](#)**person**

NOUN

an individual human, usually an adult. The plural is people, but in formal or official language the form persons is used

Synonyms or related words for this meaning of person:

## General words for person or people

**person** NOUN

an individual human, usually an adult. The plural is people, but in formal or official language the form persons is used

**people** NOUN

used for referring to humans in general

**human** NOUN

a person

**individual** NOUN

a person considered separately from their society or community

**humanity** NOUN

all the people who are living in the world

**folk** NOUN

people in general

**society** NOUN

people in general living together in organized communities, with laws and traditions controlling the way that they behave towards one another

**man** NOUN

people in general

**the many** ADVERB, DETERMINER, PREDETERMINER, PRONOUN

the majority of people

**your fellow man/men**

other people

Figur 4. Tesaurusvisning for person i Macmillans netordbog

resultatet derudfra. Hvis der fx blandt *person*'s næsten 3.600 underbegreber er 200 der alle er beskrevet med relationen ER-MEDLEM, og hvis der i alle tilfældene er tale om ord der beskrives som hyponymer til *land*, så kan dette automatisk ses som en vigtig gruppe ('landes indbyggerbetegnelser'). Men tildelingen af passende overskrifter til de grupper der kan findes på denne måde, kan næppe foretages på anden måde end ved manuel redigering.

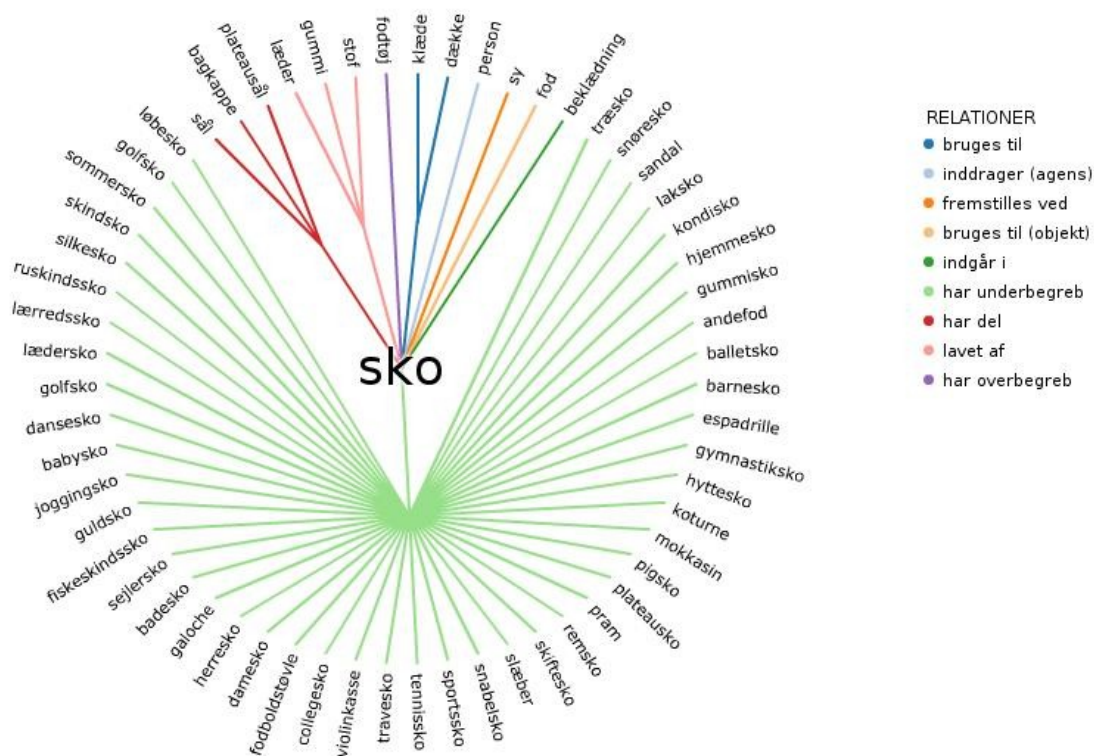
### 4.3. Visualisering og navigation

Indtil nu har vi især beskæftiget os med brugen af beslægtede ord til tekstproduktion. Men også den pædagogiske funktion fortjener at blive omtalt. Listen med beslægtede ord vil forhåbentlig inspirere brugere med kognitive behov til at klikke på nogle af ordene og på den måde gå på opdagelse i de semantiske felter. Man kan faktisk godt komme grundigt omkring i et semantisk felt fordi det jo for hver betydning oplyses hvilke beslægtede ord der er over-, under- og søsterbegreber.

#### sko

SUBSTANTIV

fodbeklædning som består af en sål (med el. uden h ...



Figur 5. Visuel visning af sko på <http://andreord.dk>

Men ellers er listeformen ikke den mest velegnede til at skabe et samlet overblik over et betydningsfelt. Planerne har derfor snarere koncentreret sig om en præsentation hvor man visuelt kan se et ords betydningsrelationer til andre ord, svarende til *The Visual Thesaurus*' af indholdet i Princeton WordNet. Men som det nogle gange går når man stiller resurser frit til rådighed som open source, kan man blive overhalet af andre, og sådan er det gået i dette tilfælde. En bruger, Anders Johannsen, har udviklet en grafisk grænseflade, [andreord.dk](http://andreord.dk), der viser DanNet-data i netop sådan en form, se et eksempel i figur 5.

Resultatet er blevet så godt at vi ikke ser nogen grund til at udvikle en tilsvarende funktion selv. I stedet bringer vi et link til denne side fra funktionen "Beslægtede ord". Siden er dynamisk på den måde at man med et klik på et af de beslægtede ord kan bringe dette ord i centrum for visningen og se dets relationer til andre ord. Ud over den visuelle præsentation er der en selvstændig søgefunktion, og i tilknytning til den visuelle præsentation bringes eksempelsætninger, en oversigt i listeform over underbegreberne med uddrag af definitionen, ligesom der oplyses om begrebets ontologiske type.

Foruden denne visning har vi gjort os overvejelser om en visning i en slags træstruktur som dels giver mulighed for en mere differentieret præsentation, dels er bedre egnet til navigation og visualisering af det samlede begrebshierarki. Figur 6 viser en ikke-levende prototype af hvordan det kunne se ud.

Det kan ses at brugeren her selv kan indstille hvordan han eller hun ønsker resultatet grupperet. I eksemplet er valgt "formål", men i andre tilfælde kan andre kriterier være mere oplagte. Der vises kun nærmeste overbegreb og underbegreb, men også dette er dynamisk sådan at siden vil positionere sig med det valgte begreb i centrum hvis brugeren klikker på enten overbegrebet (i figur 6: *motorkøretøj*) eller et af underbegreberne (fx *ambulance*). Uden for billedet er der desuden mulighed for at se hele overbegrebshierarkiet samlet og på den måde få et bedre samlet overblik over det semantiske felt. Pga. de mange forgreninger der kan være nedadtil, er det derimod ikke muligt at give et tilsvarende overblik over underbegreberne.

I skrivende stund er det ikke afklaret om strukturvisningen skal tilbydes som en selvstændig søgefunktion med egen tilgang, eller om den bedre lader sig integrere i artiklen for den søgte betydning, i tilknytning til "Beslægtede ord". Det er en af de udfordringer som redaktionen vil tage fat på i den videre udvikling af funktionen.

[ automobil 1 | **bil 1** | dyt 2.1 | karet 1.2 | slæde 1.3 | vogn 2 ]

**DEFINITION** firehjulet motorkøretøj til person- el. godstransport

**ORD MED DENNE BETYDNING**

- ▶ automobil (gammeldags; især formelt el. spøgende)
- ▶ bil
- ▶ dyt (slang)
- ▶ karet (slang)
- ▶ slæde (slang)
- ▶ vogn (især gammeldags)

**Betydningstræk**

Vis mere +

**Begrebshierarki**



Figur 6: Prototype af en mulig strukturvisning i DDO for bil

## LITTERATUR

- Andersen, Harry (1945). Dansk Begrebsordbog. København: Munksgaard.
- Bergenholtz, Henning & Vibeke Vrang 2005. Den Danske Ordbog bind 2 (E–H) og 3 (I–L) – en ordbog for folket eller for akademikere? I: LexicoNordica 12. Oslo: Nordisk Forening for Leksikografi. S. 169–191.
- Bergenholtz, Henning & Vibeke Vrang (2006). Danske Ordbog: en ordbog for lingvister! I: LexicoNordica 13. Oslo: Nordisk Forening for Leksikografi. S. 185–196.
- Bring, Sven Casper (1962). Svenskt ordförråd ordnat i begreppsklasser. Natur och Kultur.
- DanNet. <http://wordnet.dk>.
- Den Danske Netordbog. <http://www.ordbogen.com>.
- Den Danske Ordbog. <http://ordnet.dk/ddo>.
- Dornseiff, Franz (2004). Der deutsche Wortschatz nach Sachgruppen. 8. Auflage. Berlin, New York: de Gruyter.
- EuroWordNet. <http://www.illc.uva.nl/EuroWordNet>.
- Johanssen, Anders. Andre ord. <http://andreord.dk>.
- Macmillan Dictionary Online. <http://www.macmillandictionary.com>.
- Pedersen, Bolette Sandford, Sanni Nimb, Jørg Asmussen, Nicolai Hartvig Sørensen, Lars Trap-Jensen & Henrik Lorentzen (2009). DanNet: the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary. I: Language Resources and Evaluation, Volume 43, Number 3. Springer Netherlands. S. 269–299.
- Princeton WordNet. <http://wordnet.princeton.edu>.
- Roget, Peter Mark (1998). Roget's Thesaurus of English Words and Phrases. 1998-udgaven ved Betty Kirkpatrick. London: Penguin Books.
- Rosbach, Johan Hammond (2001). Ord og begreper – Norsk tesaurus. Oslo: Pax.
- The Visual Thesaurus. <http://www.visualthesaurus.com>.
- Trap-Jensen, Lars (2005a). Virtuelle perspektiver for ordbogsredigering: muligheder, strategier og virkelighedens begrænsning. I: LexicoNordica 12. Oslo: Nordisk Forening for Leksikografi. S. 109–121.
- Trap-Jensen, Lars (2005b). Kommentar til Henning Bergenholtz og Vibeke Vrang: Den Danske Ordbog bind 2 (E–H) og 3 (I–L) – en ordbog for folket eller for akademikere? I: LexicoNordica 12. Oslo: Nordisk Forening for Leksikografi. S. 189–198.



